

Pavel Diordits

Lead Data Engineer

pavel.diordits@yandex.ru · t.me/PavelDiordits · github.com/DiorditsPV · paveldiordits.site

HIGHLIGHTS

- **6+ years in data.** Lead engineer since 2022, across retail, banking and public sector
- **Pipelines.** Airflow · Spark · Kafka · dbt, batch and near-realtime
- **Production ownership.** 24/7 on-call, under one P1 incident per month
- **High-load services.** Go services up to 3000 rps, 100% test coverage
- **DWH & Data Lake.** Data Vault modelling, marts, data catalog publication

EXPERIENCE

Lead Data Engineer

Jun 2025 — present

X5 Tech

Data platform behind demand forecasting and replenishment planning: pipelines from source ingestion through distributed computation to business-facing marts.

Python · Airflow · Spark · Trino · Iceberg · PostgreSQL · Kubernetes · S3

Lead Data Engineer / Golang Developer

Dec 2022 — Jun 2025

Ozon Tech

Replaced the streaming stack with Go services that rebuild the sales model straight from core-system data, removing the team's dependency on other teams' feeds: Kafka consumers with gRPC enrichment, up to 3000 rps at 100% test coverage, rated high-critical and held under one P1 incident a month. Set up the team's dbt and Airflow development process with CI/CD across staging and production, and shipped shared ETL and Go client libraries. On-call 24/7 for first-priority incidents.

Go · Python · Kafka · gRPC · Airflow · dbt · HDFS · PostgreSQL · Redis

Data Engineer

Sep 2021 — Dec 2022

Sber

Built the warehouse behind marketing analytics and customer communications across the ecosystem's products, and migrated legacy Teradata and SAP processes onto a Hadoop stack. Delivered customer scoring and segmentation marts published to the corporate data catalog, with documentation, tests and ongoing support.

Scala · Spark · Hadoop · Oozie · Teradata · SAP · SQL

Data Engineer

Mar 2020 — Sep 2021

Moscow Department of Information Technology

Designed a Data Vault warehouse and wrote the library that integrates sources and routes data across its components. Built the publication process for the datamart layer in the data catalog and delivered interagency data transport and enrichment projects.

Python · Kafka · Airflow · Data Vault · SQL

EDUCATION

BSc, Supply Chain Management

2018

MADI — Moscow Automobile and Road State Technical University

PROJECTS

claude-skills — [Skill marketplace where the data never leaks](#)

A plugin marketplace of personal agent skills, built on one rule: the code is public, the data a skill accumulates never is. Two branches carry the split — public code on the main branch, working data pushed only to a private remote — and git hooks enforce it mechanically, so the boundary holds without relying on discipline.

interview-graph — [The whole interview on one board](#)

Questions for the data engineering stack laid out on a canvas — a column per track, ranked by difficulty inside it. The interviewer scores from the keyboard and watches coverage build up; the session ends as a self-contained HTML report on the candidate.

loop-feature — [One feature per branch, start to finish](#)

A framework that walks an agent through designing, building, verifying and shipping a feature — one branch at a time. Drops into any repo: state lives in a markdown store, code on feature branches.

ClawdIcons — [VS Code icon theme](#)

File and folder icons for Claude Code artifacts, merged on top of whatever icon theme is already in use, with path-aware badges — only the paths the tool actually reads from get painted.

book-to-anki — [Books turned into vocabulary decks](#)

Extracts rare words from FB2 and EPUB books and builds an Anki deck out of them: everything already in the Oxford 3000 is dropped, the rest is counted by frequency, translated and packed into a deck.

STACK

Python · Airflow · Spark · Trino · Iceberg · PostgreSQL · Kubernetes · S3 · Go · Kafka · gRPC · dbt · HDFS · Redis · Scala · Hadoop · Oozie · Teradata · SAP · SQL · Data Vault